DOCUMENT RESUME

ED 430 024 TM 029 737

AUTHOR Guskey, Thomas R.
TITLE New Perspectives on Evaluating Professional Development.
PUB DATE 1999-04-00
NOTE 25p.; Paper presented at the Annual Meeting of the American Educational Research Association (Montreal, Quebec, Canada, April 19-23, 1999).
PUB TYPE Reports - Descriptive (141) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Data Analysis; Data Collection; *Evaluation Methods; *Professional Development; *Research Methodology

ABSTRACT

Basic questions about the evaluation of professional development efforts are explored, including the nature and purposes of evaluation, the critical levels of professional development evaluation, and the difference between evidence and proof in evaluation. Evaluation, which is defined as the systematic investigation of merit or worth, can be characterized as planning, formative, or summative evaluation. All three types of evaluation involve the collection and analysis of data. In evaluating professional development, there are five critical levels of information to consider. These are: (1) participants' reactions; (2) participants' learning; (3) organization support and change; (4) participants' use of new knowledge and skills; and (5) student learning outcomes. In the real-world setting of professional development evaluation, it is nearly impossible to obtain proof of the impact of the effort, but it is possible to obtain good evidence. A list of guidelines is included to help improve the quality of professional development evaluations. (Contains 1 figure and 25 references.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 430 024

New Perspectives on Evaluating Professional Development

Thomas R. Guskey

University of Kentucky

Send correspondence to:

Thomas R. Guskey
College of Education
University of Kentucky
Lexington, KY 40506
Phone: 606/257-8666

E-mail: GUSKEY@POP.UKY.EDU

BEST COPY AVAILABLE

This paper was presented at the annual meeting of the
American Educational Research Association, Montreal, Canada,
April 1999.

TM029737

PERMISSION TO REPRODUCE AND
DISSEMINATE THIS MATERIAL
HAS BEEN GRANTED BY

Thomas Guskey

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

☐ This document has been reproduced as
received from the person or organization
originating it.

☐ Minor changes have been made to
improve reproduction quality.

• Points of view or opinions stated in this
document do not necessarily represent
official OERI position or policy.

New Perspectives on Evaluating Professional Development

For many years educators have operated under the premise that professional development is good by definition and, therefore, more is always better. If you want to improve your professional development program, simply add a day or two.

Today, however, we live in an age of accountability. Students are expected to meet higher standards, teachers are held accountable for student results, and professional developers are asked to show that what they do really matters. For many, this is a scary situation. They live in fear that a new superintendent or board member will come in who wants to know about the payoff from the district's investment in professional development. If the answers are not there, heads may roll and programs may get axed.

Historically, professional developers haven't paid much attention to evaluation. Many consider it a costly, time-consuming process that diverts attention from important planning, implementation, and follow-up activities. Others believe they simply lack the skill and expertise to become involved in rigorous evaluations. The result is that they either neglect evaluation issues completely, or leave them to "evaluation experts" who are called in at the end and asked to determine if what was done made any difference. The results of such an inadvertent process are seldom very useful.

Good evaluations are the product of thoughtful planning, the ability to ask good questions, and a basic understanding about how to find valid answers. In many ways they are simply the refinement of everyday thinking. Good evaluations provide information that is sound, meaningful, and sufficiently reliable to use in making thoughtful and responsible decisions about professional development processes and effects (Guskey & Sparks, 1991).

In this paper we consider four basic questions regarding professional development evaluation: (1) What is evaluation? (2) What are the purposes of evaluation? (3) What are the critical levels of professional development evaluation? and (4) What is the difference between evidence and proof? We conclude with a list of the guidelines for evaluating the wide range of professional development programs and activities used in schools today.

What Is Evaluation?

Just as there are many forms of professional development, there are also many forms of evaluation. In fact, each of us engages in hundreds of evaluation acts every day. We evaluate the temperature of our shower in the morning, the taste of our breakfast, the chances of rain and the need for an umbrella when we go outdoors, and the likelihood we will accomplish what we set out to do on any particular day. These everyday acts require the examination of evidence and the application of judgment. As such, each represents a form of evaluation.

The kind of evaluation in which we are interested, however, goes beyond these informal evaluation acts. Our interest is in evaluations that are more formal and systematic. While not everyone agrees on the best definition of this kind of evaluation, for our purposes, a useful operational definition is the following:

Evaluation is the systematic investigation of merit or worth.[*]

Let's take a careful look at this definition. By using the word "systematic," we are distinguishing this process from the multitude of informal evaluation acts in which we consciously or unconsciously engage. "Systematic" implies that evaluation in this context is a thoughtful,

[*] This definition is adapted from *The Program Evaluation Standards* (2nd ed.), by the Joint Committee on Standards for Educational Evaluation (1994).

intentional, and purposeful process. It is done for clear reasons and with explicit intent. Although the specific purpose of evaluation may vary from one setting to another, all good evaluations are deliberate and systematic.

"Investigation" refers to the collection and analysis of appropriate and pertinent information. While no evaluation can be completely objective, the process is not based on opinion or conjecture. It is, instead, based on the acquisition of specific, relevant, and valid evidence examined through appropriate methods and techniques.

The use of "merit or worth" in our definition implies appraisal and judgment. Evaluations are designed to determine the value of something. They help answer questions such as "Is this program or activity leading to the results that were intended? Is it better than what was done in the past? Is it better than another, competing activity? Is it worth the costs? The answers to these questions require more than a statement of findings. They demand an appraisal of quality and judgments of value, based on the best evidence available.

What Are The Purposes Of Evaluation?

The purposes of evaluation are generally classified in three broad categories, from which stem the three major types of evaluation. Most evaluations are actually designed to fulfill all three of these purposes, although the emphasis on each changes during various stages of the evaluation process. Because of this inherent blending of purposes, distinctions between the different types of evaluation are sometimes blurred. Still, differentiating their intent helps in clarifying our understanding of evaluation procedures (Stevens, Lawrenz, & Sharp, 1995). The three major types of evaluation include planning, formative, and summative evaluation.

Planning Evaluation

Planning evaluation takes place before a program or activity actually begins, although certain aspects may be continual and ongoing. It is designed to give those involved in program development and implementation a precise understanding of what is to be accomplished, what procedures will be used, and how success will be determined. In essence, it lays the groundwork for all other evaluation activities.

Planning evaluation involves appraisal, usually on the basis of previously established standards, of a program or activity's critical attributes. These include the specified goals, the proposal or plan to achieve those goals, the concept or theory underlying the proposal, the overall evaluation plan, and the likelihood that plan can be carried out with the time and resources available. It typically includes a determination of needs, assessment of the characteristics of participants, careful analysis of the context, and the collection of pertinent baseline information.

Evaluation for planning purposes is sometimes referred to as "preformative evaluation" (Scriven, 1991) and may be thought of as "preventative evaluation." It helps identify and remedy early on the difficulties that might plague later evaluation efforts. Planning evaluation also helps ensure that other evaluation purposes can be met in an efficient and timely manner.

Formative Evaluation

Formative evaluation occurs during the operation of a program or activity. Its purpose is to provide those responsible for the program with ongoing information about whether things are going as planned and whether expected progress is being made. If not, this same information can be used to guide necessary improvements (Scriven, 1967).

The most useful formative evaluations focus on the conditions for success. They address issues such as: What conditions are necessary for success? Have they been met? Can they be improved? In many cases, formative evaluation is a recurring process that takes place at multiple times throughout the life of the program or activity. Many program developers, in fact, are constantly engaged in the process of formative evaluation. The evidence they gather at each step of development and implementation usually stays in-house, but is used to make adjustments, modifications, or revisions (Worthen & Sanders, 1989).

To keep formative evaluations efficient and avoid expectations that will be disappointed, Scriven (1991) recommends using them as "early warning" evaluations. In other words, use formative evaluations as an early version of the final, overall evaluation. As development and implementation proceed, formative evaluation can consider intermediate benchmarks of success to determine what is working as expected and what difficulties must be overcome. Flaws can be identified and weaknesses located in time to make the adaptations necessary for success.

Summative Evaluation

Summative evaluation is conducted at the completion of a program or activity. Its purpose is to provide program developers and decision makers with judgments about the program's overall merit or worth. Summative evaluation describes what was accomplished, what were the consequences (positive and negative), what were the final results (intended and unintended), and, in some cases, did benefits justify the costs.

Unlike formative evaluations that are used to guide improvements, summative evaluations present decision makers with information they need to make crucial decisions about the life of a program or activity. Should it be continued? Continued with modifications? Expanded? Or

discontinued?. Ultimately, its focus is "the bottom line." Perhaps the best description of the distinction between formative and summative evaluation is one offered by Robert Stake: "When the cook tastes the soup, that's formative; when the guests taste the soup, that's summative" (quoted in Scriven, 1991, p. 169).

Unfortunately, many educators associate evaluation with its summative purposes only. Important information that could help guide planning, development, and implementation is often neglected, even though such information can be key in determining a program or activity's overall success. Summative evaluation, although necessary, often comes too late to be much help. Thus, while the relative emphasis on planning, formative, and summative evaluation changes through the life of a program or activity, all three are essential to a meaningful evaluation.

What Are The Critical Levels Of Professional Development Evaluation?

Planning, formative, and summative evaluation all involve the collection and analysis of information. In evaluating professional development, there are five critical stages or levels of information to consider. These levels represent an adaptation of an evaluation model developed by Kirkpatrick (1959) for judging the value of supervisory training programs in business and industry. Kirkpatrick's model, although widely applied, has seen limited use in education because of inadequate explanatory power. It is helpful in addressing a broad range of "what" questions, but lacking when it comes to explaining "why" (Alliger & Janak, 1989; Holton, 1996). The model presented here is designed to resolve that inadequacy.

The five levels in the model are hierarchically arranged from simple to more complex.. With each succeeding level, the process of gathering evaluation information is likely to require

more time and resources. More importantly, each higher level builds on the ones that came before. In other words, success at one level is necessary for success at the levels that follow.

Below is a brief description of each of the five levels and its importance in the evaluation process. Included are the crucial questions addressed at each level, how that information can be gathered, what is being measured, and how that information will be used. A summary of these issues is also presented in Figure 1.

Level 1: Participants' Reactions

The first level of professional development evaluation is participants' reactions to the experience. This is the most common form of professional development evaluation, the simplest, and the level at which educators have the most experience. It is also the easiest type of information to gather and analyze.

The questions addressed at this level focus on whether or not participants liked it. When they walked out, did they feel their time was well spent? Did the material make sense to them? Were the activities meaningful? Was the instructor knowledgeable and helpful? Do they believe what they learned will be helpful? Also important are questions such as, Was the coffee hot and ready on time? Were the refreshments fresh and tasty? Was the room the right temperature? Were the chairs comfortable? To some, questions such as these may seem silly and inconsequential. But experienced professional developers know the importance of attending to these basic human needs.

Information on participants' reactions is generally gathered through questionnaires handed out at the end of a session or activity. These questionnaires typically include a combination of rating-scale items and open-ended response questions that allow participants to provide more

personalized comments. Because of the general nature of this information, the same questionnaire often is used for a broad range of professional development experiences. Many professional organizations, for example, use the same questionnaire for all their professional development activities.

Measures of participants' reactions are sometimes referred to as "happiness quotients" by those who insist they measure only the entertainment value of an activity, not its quality or worth But measuring participants' initial satisfaction with the experience provides information that can help improve the design and delivery of programs or activities in valid ways. In addition, positive reactions from participants are usually a necessary prerequisite to higher level evaluation results.

Level 2: Participants' Learning

In addition to liking it, we would also hope that participants learned something from their professional development experience. Level 2 focuses on measuring the knowledge, skills, and perhaps attitudes participants gained. Depending on the goals of the program or activity, this can involve anything from a pencil-and-paper assessment (Can participants describe the critical attributes of mastery learning and give examples of how these might be applied in common classroom situations?) to a simulation or full-scale skill demonstration (Presented with a variety of classroom conflicts, can participants diagnose each situation, and then prescribe and carry out a fair and workable solution?). Oral or written personal reflections, or examination of the portfolios participants assemble can also be used to document their learning.

Although evaluation information at Level 2 sometimes can be gathered at the completion of a session, it seldom can be accomplished with a standardized form. Measures must be based on the learning goals prescribed for that particular program or activity. This means specific criteria

and indicators of successful learning must be outlined prior to the beginning of the professional development experience. Openness to possible "unintended learnings," either positive or negative, also should be considered. If there is concern that participants may already possess the requisite knowledge and skills, some form of pre- and post-assessment may be required. Analysis of this information provides a basis for improving the content, format, and organization of the program or activities.

[Insert Figure 1]

Level 3: Organization Support and Change

At Level 3 our focus shifts to the organization and, specifically, to information on organization support and change. Organizational variables can be key to the success of any professional development effort. They also can hinder or prevent success, even when the individual aspects of professional development are done right (Sparks, 1996a).

Suppose, for example, a group of educators participate in a professional development program on cooperative learning, gain a thorough understanding of the theory, and organize a variety of classroom activities based on cooperative learning principles. Following their training they try to implement these activities in schools where students are generally graded "on the curve," according to their relative standing among classmates, and great importance is attached to selecting the class valedictorian. Organizational policies and practices such as these make learning highly competitive and will thwart the most valiant efforts to have students cooperate and help each other learn (Guskey, 1996).

The lack of positive results in this case is not due to poor training or inadequate learning. Rather, it is due to organizational policies that are incompatible with implementation efforts. The

Figure I. Five Levels of Professional Development Evaluation

Evaluation Level	What Questions Are Addressed?	How Will Information Be Gathered?	What is Measured or Assessed?	How Will Information Be Used?
1. Participants' Reactions	• Did they like it? • Was their time well spent? • Did the material make sense? • Will it be useful? • Was the leader knowledgeable and helpful? • Were the refreshments fresh and tasty? • Was the room the right temperature? • Were the chairs comfortable?	• Questionnaires administered at the end of the session.	• Initial satisfaction with the experience	• To improve program design and delivery
2. Participants' Learning	• Did participants acquire the intended knowledge and skills?	• Paper-and-pencil instruments • Simulations • Demonstrations • Participant reflections (oral and/or written) • Participant portfolios	• New knowledge and skills of participants	• To improve program content, format, and organization
3. Organization Support & Change	• What was the impact on the organization? • Did it affect organizational climate and procedures? • Was implementation advocated, facilitated, and supported? • Was the support public and overt? • Were problems addressed quickly and efficiently? • Were sufficient resources made available? • Were successes recognized and shared?	• District and school records • Minutes from follow-up meetings. • Questionnaires • Structured interviews with participants and district or school administrators • Participant portfolios	• The organization's advocacy, support, accommodation, facilitation, and recognition.	• To document and improve organizational support • To inform future change efforts
4. Participants' Use of New Knowledge and Skills	• Did participants effectively apply the new knowledge and skills?	• Questionnaires • Structured interviews with participants and their supervisors • Participant reflections (oral and/or written) • Participant portfolios • Direct observations • Video or audio tapes	• Degree and quality of implementation	• To document and improve the implementation of program content
5. Student Learning Outcomes	• What was the impact on students? • Did it affect student performance or achievement? • Did it influence students' physical or emotional well-being? • Are students more confident as learners? • Is student attendance improving? • Are dropouts decreasing?	• Student records • School records • Questionnaires • Structured interviews with students, parents, teachers, and/or administrators • Participant portfolios	• Student learning outcomes: • Cognitive (Performance & Achievement) • Affective (Attitudes & Dispositions) • Psychomotor (Skills & Behaviors)	• To focus and improve all aspects of program design, implementation, and follow-up • To demonstrate the overall impact of professional development

BEST COPY AVAILABLE

gains made at Levels 1 and 2 are essentially canceled by problems at Level 3 (Sparks & Hirsh, 1997). That is why it is essential to gather information on organization support and change.

Questions at Level 3 focus on the organizational characteristics and attributes necessary for success. Was the advocated change aligned with the mission of the organization? Was change at the individual level encouraged and supported at all levels? Did the program or activity affect organizational climate and procedures? Was administrative support public and overt? Were problems addressed quickly and efficiently? Were sufficient resources made available, including time for sharing and reflection (Langer & Colton, 1994)? Were successes recognized and shared? Issues such as these can be major contributing factors to the success of any professional development effort.

Gathering information on organization support and change is generally more complicated than previous levels. Procedures also differ depending on the goals of the program or activity. They may involve analyses of district or school records, or examination of the minutes from follow-up meetings. Questionnaires sometimes can be used to tap issues such as the organization's advocacy, support, accommodation, facilitation, and recognition of change efforts. Structured interviews with participants and district or school administrators can be helpful as well. This information is used not only to document and improve organizational support, but also to inform future change initiatives.

Level 4: Participants' Use of New Knowledge and Skills

With organizational variables set aside, we turn our attention to whether participants are using their new knowledge and skills on the job. At Level 4 our central question is, "Did what participants' learn make a difference in their professional practice?" The key to gathering relevant

information at this level rests in the clear specification of indicators that reveal both the degree and quality of implementation. In other words, how can you tell if what participants learned is being used and being used well? Depending on the goals of the program or activity, this may involve questionnaires or structured interviews with participants and their supervisors. Oral or written personal reflections, or examination of participants' journals or portfolios also can be considered. The most accurate information is likely to come from direct observations, either with trained observers or by reviewing video or audio tapes. When observations are used, however, they should be kept as unobtrusive as possible (for examples, see Hall & Hord, 1987).

Unlike Levels 1 and 2, information at Level 4 cannot be gathered at the completion of a professional development session. Measures of use must be made after sufficient time has passed to allow participants to adapt the new ideas and practices to their setting. Because implementation is often a gradual and uneven process, measures also may be necessary at several time intervals. This is especially true if there is interest in continuing or on-going use. Analysis of this information provides evidence on current levels of use and can help restructure future programs and activities to facilitate better and more consistent implementation.

Level 5: Student Learning Outcomes

At Level 5 we address what is typically "the bottom line" in education: What was the impact on students? Did the professional development program or activity benefit students in any way? The particular outcomes of interest will depend, of course, on the goals of that specific professional development effort. In addition to the stated goals, certain "unintended" outcomes may be important as well. For this reason, multiple measures of student learning are always essential at Level 5 (Joyce, 1993).

Consider the example of a group of elementary educators who devote their professional development time to finding ways to improve the quality of students' writing. In a study group they explore the research on writing instruction, analyze various approaches, and devise a series of strategies they believe will work for their students. In gathering Level 5 information, they find students' scores on measures of writing ability increased significantly over the course of the school year when compared to the progress of comparable students who were not involved in these strategies. On further analysis, however, they discover that over the same time period, their students' scores on measures of mathematics achievement declined. This "unintended" outcome apparently occurred because instructional time in mathematics was inadvertently sacrificed to provide more time for students to work on their writing. Had information at Level 5 been restricted to a single measure of students' writing, this important "unintended" result would not have been identified.

Measures of student learning typically include indicators of student performance and achievement, such as assessment results, portfolio evaluations, marks or grades, and scores from standardized examinations. But in addition to these cognitive indicators, affective (attitudes and dispositions) and psychomotor outcomes (skills and behaviors) may be considered as well. Examples include assessments of students' self-concepts, study habits, school attendance, homework completion rates, or classroom behaviors. Schoolwide indicators such as enrollment in advanced classes, memberships in honor societies, participation in school-related activities, disciplinary actions, and retention or drop-out rates might also be considered.

The major source of such information is student and school records. Results from questionnaires and structured interviews with students, parents, teachers, and/or administrators could also be included. The summative purpose of this information is to document a program or

activity's overall impact. But formatively, it can be used to inform improvements in all aspects of professional development, including program design, implementation, and follow-up. In some cases information on student learning outcomes is used to estimate the cost effectiveness of professional development, or what is sometimes referred to as "return on investment," or "ROI evaluation" (Parry 1996; Todnem & Warner, 1993).

Evaluation at any of these five levels can be done well or poorly, convincingly or laughably. The information gathered at each level is important and can help improve professional development programs and activities. But as many have discovered, tracking efficiency at one level tells you nothing about effectiveness at the next. Although success at an early level may be necessary for positive results at the next higher one, it is clearly not sufficient. That is why each level is important. Sadly, the bulk of professional development today is evaluated only at Level 1, if at all. Of the rest, the majority are measured only at Level 2 (Cody & Guskey, 1997).

What Is The Difference Between Evidence And Proof?

Now that you know about planning, formative, and summative evaluation, and understand the five levels involved in evaluating professional development, are you ready to "prove" that your professional development programs make a difference? With this new knowledge can you demonstrate that what was done in professional development, and nothing else, is solely responsible for that ten percent increase in student achievement scores? For the five percent decrease in dropout rate? For the 50 percent reduction in recommendations to the office for disciplinary action?

Are you trying to say the counseling department had nothing to do with it? Do the principal and assistant principal get no credit for their support and encouragement? Might not year-to-year fluctuations in students have something to do with the results? And consider the other side of the coin. If achievement ever happens to drop following some highly touted professional development initiative, would you be willing to accept full blame for the loss?

Arguments about whether you can absolutely, positively isolate the impact of professional development on improvements in student performance are generally irrelevant. In most cases, you simply cannot get ironclad proof (Kirkpatrick, 1977). To do so you would need to eliminate or control for all other factors that could have caused the change. This requires the random assignment of educators and students to experimental and control groups. The experimental group would take part in the professional development activity while the control group would not. Comparable measures would then be gathered from each and the differences tested.

The problem, of course, is that nearly all professional development takes place in real-world settings where such experimental conditions are impossible to meet. The relationship between professional development and improvements in student learning in these real-world settings is far too complex and there are too many intervening variables to allow for simple causal inferences (Guskey, 1997; Guskey & Sparks, 1996). What's more, most schools are engaged in systemic reform initiatives that involve the simultaneous implementation of multiple innovations (Fullan, 1992). Isolating the effects of a single program or activity under such conditions is usually impossible.

But in the absence of proof, you can collect awfully good "evidence" about whether or not professional development is contributing to specific gains in student learning. Setting up meaningful comparison groups and using appropriate pre- and post-measures provides extremely

valuable information. Time-series designs that include multiple measures collected before and after implementation are another useful alternative. Above all, you must be sure to gather evidence on measures that are meaningful to stakeholders in the evaluation process. Evidence is what most people want anyway. Superintendents and board members rarely ask, "Can you prove it?" What they ask for is evidence.

Consider, for example, the use of anecdotes and testimonials. From a methodological perspective, they are a poor source of data. They are typically biased and highly subjective. They may be inconsistent and unreliable. Nevertheless, they are a personalized form of information that can be powerful and convincing. And as any trial attorney will tell you, they offer the kind of evidence that most people believe. Although it would be imprudent to base your entire evaluation on anecdotes and testimonials, they are an important source of evidence that should never be ignored.

Keep in mind, too, that good evidence is not that hard to come by if you know what you're looking for before you begin. If you do a good job of clarifying your goals up front, most evaluation issues pretty much fall into line. The reason many educators think evaluation at Levels 4 and 5 is so difficult, expensive, and time-consuming, is because they are coming in after the fact to search for results. It is as if they are saying, "We don't know what we are doing or why we are doing it, but let's find out if anything happened" (Gordon, 1991). If you don't know where you are going, it's very difficult to tell if you've arrived.

So when it comes to evidence versus proof, the message is this: *Always seek proof, but collect lots of evidence along the way.* Because of the nature of most professional development efforts, your evidence may be more exploratory than confirmatory. Still, it can offer important indications about whether you are heading in the right direction or whether you need to go back

to the drawing board. Remember, too, that knowing ahead of time what you are trying to accomplish will make it much easier to identify the kind of evidence you need.

Evaluation Guidelines

It should be clear by now that good evaluations of professional development don't have to be costly. Nor do they demand sophisticated technical skills, although technical assistance can sometimes be helpful. What they do require is the ability to ask good questions and a basic understanding about how to find valid answers. Good evaluations provide sound, useful, and sufficiently reliable information that can be used to make thoughtful and responsible decisions about professional development processes and effects.

Following is a list of guidelines designed to help improve the quality of professional development evaluations. Although strictly adhering to these guidelines won't guarantee your evaluation efforts will be flawless, it will go a long way toward making them more meaningful, more useful, and far more effective.

Planning Guidelines

1. *Clarify the intended goals.* The first step in any evaluation is to make sure your professional development goals are clear, especially in terms of the results you hope to attain with students and the classroom or school practices you believe will lead to those results. Change experts refer to this as "Beginning with the end in mind." It is also the premise of a "results-driven" approach to professional development (Sparks, 1995, 1996b).

2. *Assess the value of the goals.* Take steps to ensure the goals are sufficiently challenging, worthwhile, and considered important by all those involved in the professional development

process. Broad-based involvement at this stage contributes greatly to a sense of shared purpose and mutual understanding. Clarifying the relationship between established goals and the school's mission is a good place to begin.

3. *Analyze the context.* Identify the critical elements of the context where change is to be implemented and assess how these might influence implementation. Such an analysis might include the examination of pertinent baseline information on students' and teachers' needs, their unique characteristics and background experiences, the resources available, the level of parent involvement and support, and the organizational climate.

4. *Estimate the program's potential to meet the goals.* Explore the research base of the program or activity, and the validity of the evidence supporting its implementation in contexts similar to yours. When exploring the literature on a particular program, be sure to distinguish facts from persuasively argued opinions. A thorough analysis of the costs of implementation, and what other services or activities must be sacrificed to meet those costs, should be included as well.

5. *Determine how the goals can be assessed.* Decide, up front, what evidence you would trust in determining if the goals are attained. Ensure that evidence is appropriate, relevant to the various stakeholders, and meets at least minimal requirements for reliability and validity. Keep in mind, too, that multiple indicators are likely to be necessary in order to tap both intended and possible unintended consequences.

6. *Outline strategies for gathering evidence.* Determine how that evidence will be gathered, who will gather it, and when it should be collected. Be mindful of the critical importance of intermediate or benchmark indicators that might be used to identify problems (formative) or forecast final results (summative). Select procedures that are thorough and systematic, but considerate of participants' time and energy. Thoughtful evaluations typically use a combination

of quantitative and qualitative methods, based on the nature of the evidence sought. To document improvements you must also plan meaningful contrasts with appropriate comparison groups, pre- and post-measures, or longitudinal time-series measures.

Formative and Summative Guidelines

7. *Gather and analyze evidence on participants' reactions.* At the completion of both structured and informal professional development activities, collect information on how participants regard the experience. A combination of items or methods is usually required to assess perceptions of various aspects of the experience. In addition, keeping the information anonymous generally guarantees more honest responses.

8. *Gather and analyze evidence on participants' learning.* Develop specific indicators of successful learning, select or construct instruments or situations in which that learning can be demonstrated, and collect the information through appropriate methods. The methods used will depend, of course, on the nature of the learning sought. In most cases a combination of methods or procedures will be required.

9. *Gather and analyze evidence on organization support and change.* Determine the organizational characteristics and attributes necessary for success, and what evidence best illustrates those characteristics. Then collect and analyze that information to document and to improve organizational support.

10. *Gather and analyze evidence on participants' use of new knowledge and skills.* Develop specific indicators of both the degree and quality of implementation. Then determine the best methods to collect this information, when it should be collected, and how it can be used to offer participants constructive feedback to guide (formative) or judge (summative) their implementation

efforts. If there is concern with the magnitude of change (Is this really different from what participants have been doing all along?), pre- and post-measures may need to be planned. The methods used to gather this evidence will depend, of course, on the specific characteristics of the change being implemented.

11. *Gather and analyze evidence on student learning outcomes.* Considering the procedures outlined in Step 6, collect the student information that most directly relates to the program or activity's goals. Be sure to include multiple indicators to tap the broad range of intended and possible unintended outcomes in the cognitive, affective, and psychomotor areas. Anecdotes and testimonials should be included to add richness and provide special insights. Analyses should be based on standards of desired levels of performance over all measures and should include contrasts with appropriate comparison groups, pre- and post-measures, or longitudinal time-series measures.

12. *Prepare and present evaluation reports.* Develop reports that are clear, meaningful, and comprehensible to those who will use the evaluation results. In other words, present the results in a form that can be understood by decision makers, stakeholders, program developers, and participants. Evaluation reports should be brief but thorough, and should offer practical recommendations for revision, modification, or further implementation. In some cases reports will include information comparing costs to benefits, or the "return on investment."

Conclusion

Over the years a lot of good things have been done in the name of professional development. So have a lot of rotten things. What professional developers haven't done is provide evidence to document the difference between the good and the rotten. Evaluation is the

key, not only to making those distinctions, but also to explaining how and why they occurred. To do this we must recognize the important summative purposes that evaluation serves, and its vital planning and formative purposes as well.

Just as we urge teachers to plan carefully and make ongoing assessments of student learning an integral part of the instructional process, we need to make evaluation an integral part of the professional development process. Systematically gathering and analyzing evidence to inform what we do must become a central component in professional development technology. Recognizing and using this component will tremendously enhance the success of professional development efforts everywhere.

References

Alliger, G. M., & Janak, E. A. (1989). Kirkpatrick's levels of training criteria: Thirty years later. *Personnel Psychology*, *42*(2), 331-342.

Cody, C. B., & Guskey, T. R. (1997). Professional development. In J. C. Lindle, J. M. Petrosko, & R. S. Pankratz (Eds.), *1996 Review of research on the Kentucky Education Reform Act* (pp. 191-209). Frankfort, KY: The Kentucky Institute for Education Research.

Fullan, M. G. (1992). Visions that blind. *Educational Leadership*, *49*(5), 19-20.

Gordon, J. (1991). Measuring the 'goodness' of training. *Training* (August), 19-25.

Guskey, T. R. (1996). Reporting on student learning: Lessons from the past -- Prescriptions for the future. In T. R. Guskey (Ed.), *Communicating Student Learning. 1996 Yearbook of the Association for Supervision and Curriculum Development* (pp. 13-24). Alexandria, VA: Association for Supervision and Curriculum Development.

Guskey, T. R. (1997). Research needs to link professional development and student learning. *Journal of Staff Development, 18*(2), 36-40.

Guskey, T. R., & Sparks, D. (1991). What to consider when evaluating staff development. *Educational Leadership, 49*(3), 73-76.

Guskey, T. R., & Sparks, D. (1996). Exploring the relationship between staff development and improvements in student learning. *Journal of Staff Development, 17*(4), 34-38.

Hall, G. E., & Hord, S. M. (1987). *Change in schools: Facilitating the process.* Albany, NY: SUNY Press.

Holton, E. F. (1996). The flawed four-level evaluation model. *Human Resources Development Quarterly, 7*(1), 5-21.

Johnson, B. M. (1995). Why conduct action research? *Teaching and Change, 3*(1), 90-104.

Joint Committee on Standards for Educational Evaluation (1994). *The program evaluation standards* (2nd ed.). Thousand Oaks, CA: Sage.

Joyce, B. (1993). The link is there, but where do we go from here? *Journal of Staff Development, 14*(3), 10-12.

Kirkpatrick, D. L. (1959). Techniques for evaluating training programs. A four-part series beginning in the November issue (Vol. 13, No. 11) of *Training and Development Journal* (then titled *Journal for the American Society of Training Directors*).

Kirkpatrick, D. L. (1977). Evaluating training programs: Evidence vs. proof. *Training and Development Journal, 31*(11), 9-12.

Parry, S. B. (1996). Measuring training's ROI. *Training & Development, 50*(5), 72-75.

Scriven, M. (1967). The methodology of evaluation. In R. E. Stake (Ed.), *Curriculum evaluation.* American Educational Research Association Monograph Series on Evaluation, No. 1. Chicago: Rand McNally.

Scriven, M. (1991). *Evaluation thesaurus* (4th ed.). Newbury Park, CA: Sage.

Stevens, F., Lawrenz, F. & Sharp, L. (1995). *User-friendly handbook for project evaluation: Science, mathematics, engineering, and technology education.* Arlington, VA: National Science Foundation.

Sparks, D. (1995, April). Beginning with the end in mind. *School Team Innovator, 1*(1), p. 1.

Sparks, D. (1996a, February). Viewing reform from a systems perspective. *The Developer*, pp. 2, 6.

Sparks, D. (1996b, January). Results-driven staff development. *The Developer*, p. 2.

Sparks, D., & Hirsh, S. (1997). *A new vision for staff development.* Alexandria, VA: Association for Supervision and Curriculum Development.

Todnem, G., & Warner, M. P. (1993). Using ROI to assess staff development efforts. *Journal of Staff Development, 14*(3), 32-34.

Worthen, B. R., & Sanders, J. R. (1989). *Educational evaluation.* New York: Longman.